Handout 7

STAT5710 Applied Stochastic Processes

Motivation for MCMC: a simple but intractable likelihood

In the Dirichlet-multinomial model the likelihood $L(P) = \prod_i \prod_j p_{ij}^{N_{ij}}$ is conditional on x_0. Suppose instead that the chain starts in stationarity, so that $X_0 \sim \pi = \pi(P)$.

1. Write the likelihood that results.

$$L(P) \; =$$

2. With a Dirichlet(α_i) prior on each row, write the posterior up to proportionality.

$$f(P \mid \{x_t\}) \; \propto$$

3. Write the normalizing constant that we would have to evaluate.

4. Suppose we could somehow sample $P^{(1)}, \ldots, P^{(B)} \overset{iid}{\sim} f(P \mid \{x_t\})$.

   (a) Averaging gives a consistent estimate of P:

$$\hat{p}_{ij} \; = \qquad\qquad \overset{a.s.}{\longrightarrow} \qquad\qquad \text{as}$$

   (b) Further, we could estimate any functional $\theta = g(P)$ consistently:

$$\hat{\theta} \; = \qquad\qquad \overset{a.s.}{\longrightarrow} \qquad\qquad \text{as}$$

5. Independent sampling is tricky. Suppose instead we had a Markov chain whose stationary distribution was $f(P \mid \{x_t\})$. Then one could approximate an *iid* sample by:

   i.

   ii.

   The almost sure convergence in (4) still holds provided that _____.

The Metropolis–Hastings algorithm

Let $f(x)$ be a **target** distribution, discrete or continuous, and let $q(y \mid x)$ be a **proposal** distribution.

6. Fill in the algorithm.

   Initialize X_0. For $t = 0, 1, 2, \ldots, T$:

   (a) Propose $Y \sim$

   (b) Update

$$X_{t+1} \;=\; \begin{cases} \text{with probability } \alpha(X_t, Y) \\ \\ \text{with probability } 1 - \alpha(X_t, Y) \end{cases}$$

   where the acceptance probability is

$$\alpha(x, y) \;=$$

   Return $X^{(1)}, \ldots, X^{(B)}$.

7. Usually $B \neq T$ because…

Question: why is f the stationary distribution of the chain this produces? To answer this we need to extend discrete-time Markov chain concepts to an arbitrary and possibly continuous state space S.

Markov chains on a general state space

For a discrete-time chain on a countable state space we wrote

$$P(x, y) = P(X_{t+1} = y \mid X_t = x)$$

and for any $A \subseteq S$ we have

$$P(X_{t+1} \in A \mid X_t = x) = \sum_{y \in A} P(x, y)$$

8. **Definition.** For a chain on an arbitrary state space S, the **transition kernel** is

$$K(x, A) \;= \qquad\qquad\qquad\qquad x \in S,\ A \subseteq S$$

   (a) In the continuous case, if a density exists, $\quad K(x, A) =$

   (b) In the discrete case, $\quad K(x, A) =$

9. **Definition.** π is a **stationary distribution** if

   (a) for countable S: $\pi(y) \; =$ for all $y \in S$

   (b) for uncountable S: $\pi(y) \; =$ for all $y \in S$

10. **Definition.** $\{X_n\}$ is **reversible** with respect to π if it satisfies the *detailed balance equations*

    for all $x, y \in S$

11. **Lemma.** If $\{X_n\}$ is reversible with respect to π, then π is a stationary distribution.

    Proof. Sum the detailed balance equations over x:

    $$\sum_x \pi(x)\, k(y \mid x) \; =$$

    In the continuous case, replace sums by integrals:

    $$\int_S \pi(x)\, k(y \mid x) dx \; =$$

12. **Lemma.** If $\{X_n\}$ is reversible with respect to π and $X_n \sim \pi$, then

    $$P(X_n \in A,\, X_{n+1} \in B) \; = \; P(X_n \in B,\, X_{n+1} \in A)$$

    Proof. Show the result in the discrete case. For the continuous case, replace sums by integrals.

    $$P(X_n \in A, X_{n+1} \in B) =$$

The theorem

13. **Theorem.** Provided $q(y \mid x) > 0 \implies q(x \mid y) > 0$, the chain $\{X_t\}$ constructed by Metropolis–Hastings has the target f as its stationary distribution.

    By the lemma in (11), it is enough to show that…

    Proof. Fix $x \neq y$ and write the acceptance ratio as

    $$r(x, y) \;=$$

    (a) Express both acceptance probabilities in terms of r.

    $$\alpha(x, y) =$$

    $$\alpha(y, x) =$$

    (b) For $y \neq x$, the transition density/mass of the Metropolis–Hastings chain is

    $$k(y \mid x) \;=$$

    (c) Suppose $r(x, y) \leq 1$, so that $\alpha(x, y) =$ _____ and $\alpha(y, x) =$ _____. Then f satisfies the detailed balance equations:

    $$f(x)\, k(y \mid x) \;=$$

    (d) Now suppose $r(x, y) > 1$, so that $\alpha(x, y) =$ _____ and $\alpha(y, x) =$ _____. Similarly:

    $$f(y)\, k(x \mid y) \;=$$

    So f satisfies the detailed balance equations, and by the lemma, is a stationary distribution of $\{X_t\}$.